Proc. Natl. Acad. Sci. USA
Vol. 75, No. 3, pp. 1274–1278, March 1978
Biochemistry

Spliced early mRNAs of simian virus 40

(S_1 endonuclease/exonuclease VII/transcription mapping/viable deletion mutants)

ARNOLD J. BERK AND PHILLIP A. SHARP

Department of Biology and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Communicated by J. D. Watson, January 9, 1978

ABSTRACT **Biochemical methods are presented for determining the structure of spliced RNAs present in cells at low concentrations. Two cytoplasmic spliced viral RNAs were detected in CV-1 cells during the early phase of simian virus 40 (SV40) infection. One is 2200 nucleotides in length and is composed of two parts, 330 and 1900 nucleotides, mapping from ~0.67 to ~0.60 and from ~0.54 to ~0.14, respectively, on the standard viral map. The other is 2500 nucleotides long and also is composed of two parts, 630 and 1900 nucleotides mapping from ~0.67 to ~0.54 and from ~0.54 to ~0.14, respectively. Correlation of the structure of these mRNAs with the structure of the early SV40 proteins, small T antigen (17,000 daltons) and large T antigen (90,000 daltons), determined by others suggests that: (*i*) translation of the 2500-nucleotide mRNA yields small T antigen; (*ii*) translation of the 2200-nucleotide mRNA proceeds through the splice point in the RNA to produce large T antigen (and thus large T antigen is encoded in two separate regions of the viral genome); and (*iii*) the DNA sequences between ~0.67 and ~0.60 present in both mRNAs are translated in the same reading frame in both mRNAs to yield two separate gene products that have the same NH_2-terminal sequence. Therefore, expression of the early SV40 genes is partially controlled at the level of splicing of RNAs.**

The genome of simian virus 40 (SV40) is a closed circular duplex DNA molecule approximately 5000 base pairs in length. The genetic information encoded in approximately half of this DNA is expressed during the early phase of infection in the form of two proteins (1–3). The larger of these, large T antigen, has a molecular weight estimated by sodium dodecyl sulfate/polyacrylamide gel electrophoresis to be 90,000–100,000 (4), while the other, small T antigen, has a molecular weight of approximately 17,000 (1–3). The small T antigen shares a considerable fraction of its amino acid sequence with the large T antigen, suggesting that portions of these two proteins are encoded in the same DNA sequence (2, 3, 5).

The sequences present in the early mRNAs that encode these two proteins are transcribed in the counterclockwise direction from the early region of the genome mapping between roughly 0.67 and 0.17 unit lengths from the single *Eco*RI site (6, 7) (Fig. 1). Both early proteins can be translated *in vitro* from mRNAs sedimenting at 19 S, the small T antigen mRNA sedimenting slightly faster than the large T antigen mRNA (3). These two mRNAs therefore must contain extensively overlapping sequences because each of them is approximately the size of the entire early region.

Recently, several paradoxical observations have been made concerning the organization of DNA sequences that encode the two early gene products. The basis of this paradox is that nearly the entire early region (2500–2600 nucleotides in length) would be required to encode a protein with the estimated molecular weight of large T antigen, 90,000–100,000. Yet, large deletions within the early region mapping between 0.59 and 0.54 (8) do not result in the production of an altered large T antigen (ref. 1; M. J. Sleigh, W. C. Topp, R. Hanich, and J. Sambrook, personal communication). Different examples of these deletion mutants (1) or of an independently isolated series of similar mutants (Sleigh *et al.*, personal communication) either fail to induce a detectable small T antigen or induce an altered protein that has a decrease in molecular weight roughly corresponding to the size of the deletion. Therefore, although the two early proteins are apparently encoded in partially overlapping DNA sequences, sequences mapping between 0.54 and 0.59 must code for small T antigen exclusively. In addition, DNA sequence studies (9) have shown that a transcript of the early strand would contain two termination codons in each of all three reading frames near position 0.54. In an attempt to clarify this puzzling situation, we have determined the structure of the early SV40 mRNAs by using new methods developed in our laboratory (10).

Recently it has been shown that the sequences comprising a single covalently continuous viral mRNA molecule may arise from well-separated regions on a viral genome (11–16). Such mRNA molecules are referred to as "spliced mRNAs." In this work, we find two early SV40 mRNAs that are also spliced. The map positions of the splice points and of the sequences present in these early mRNAs can account for the shared amino acid sequence of the large and small T antigens, the occurence of termination codons in all three reading frames at 0.54, the phenotype of the deletion mutants, and the relative sizes of the large and small T antigen mRNAS.

MATERIALS AND METHODS

Preparation of RNA. Subconfluent plates of CV-1 cells were infected at 37° with 5–10 plaque forming units per cell of SV40 strain 777. After 60 min, Dulbecco's modified medium containing 10% calf serum and cytosine arabinoside (20 μg/ml) was added. After 18 more hr of incubation at 37°, the cells were harvested by scraping, and cytoplasmic RNA was isolated as described (10).

Preparation of ^{32}P-Labeled DNA. Closed circular SV40 DNA labeled *in vivo* with ^{32}P had a specific activity of 1–3 × 10^6 cpm/μg and was prepared as described (17). Alternatively, SV40 DNA was labeled by nick-translation (18) as described (19) except for the following modifications: all four of the α-^{32}P labeled dNTPs were present at a specific activity of 20–40 Ci/mmol, and DNase I, 8 pg/ml, was added to the reaction which proceeded for 45 min at 15°. Approximately 30% of the dNTPs were polymerized into DNA which was >50% full-length SV40 strands. Final specific activities were 1–3 × 10^7 cpm/μg. Labeled SV40 DNA was digested to completion with the restriction endonucleases indicated in the text and figure legends. *Eco*RI, *Bam* I, and *Bgl* I were purified in our labora-

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: SV40, simian virus 40; t_m, melting temperature.

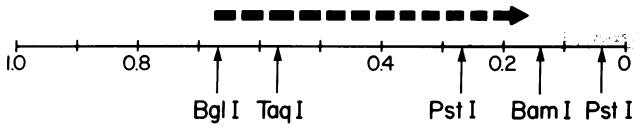

FIG. 1. Restriction map of SV40 DNA. The circular genome is represented as a line broken at the single *Eco*RI site defined as 0/1.0 and is divided into fractional unit lengths. The broken arrow represents the approximate region of the genome expressed during the early phase of infection and the direction of transcription.

tory. *Taq* I was a gift from C. Cole, and *Pst* I was purchased from New England Biolabs.

Hybridization. Restriction endonuclease-cleaved ^{32}P-labeled SV40 DNA or restriction fragments, purified as described (10), were ethanol-precipitated along with 10–20 μg of carrier yeast RNA. The pellets were briefly dried and dissolved in 20–100 μl of 80% formamide/0.4 M NaCl/0.04 M 1,4-piperazinediethanesulfonic acid (Pipes), pH 6.4/1 mM EDTA (20). Aliquots of the ethanol-precipitated early SV40 cytoplasmic RNA were pelleted and briefly air-dried. The RNA pellets were then dissolved in the hybridization buffer in which the ^{32}P-labeled SV40 DNA previously had been dissolved. Final concentrations of SV40 DNA and cytoplasmic RNA were 5 or 10 μg/ml and 2.5 to 5 mg/ml, respectively. The hybridization mixture was placed in capped 1.5-ml polypropylene tubes and incubated at 60° for 10 min and then at 49° for 3 hr, except where indicated.

Endonuclease S_1 Digestion. An aliquot of the hybridization mixture was removed and diluted into 10 volumes of 0.25 M NaCl/0.03 M NaOAc/1 mM $ZnSO_4$/5% glycerol/thermally denatured salmon sperm DNA, (20 μg/ml) chilled to 0°. S_1 purified by the method of Vogt (21) was added, 1 unit per ml of S_1 reaction mix, and the solution was incubated at 45° for 30 min. S_1-resistant material was ethanol-precipitated.

Exonuclease VII Digestion. An aliquot of the hybridization mixture was removed and diluted into 10 volumes of 0.03 M KCl/0.01 M Tris, pH 7.4/0.01 M Na_3EDTA, chilled to 0°. An amount of *Escherichia coli* exonuclease VII, purified by the method of Chase and Richardson (22), sufficient to digest 2 μg of thermally denatured linear SV40 DNA to 95% acid solubility in 60 min at 45° in this buffer was added per ml of exonuclease VII reaction mix and the solution was then incubated at 45° for 60 min. Exonuclease VII-resistant material was ethanol-precipitated. Exonuclease VII was a generous gift from S. Goff.

Gel Electrophoresis. Neutral agarose gels were 1.4% agarose in 0.04 M NaOAc/0.05 M Tris/2 mM EDTA, pH 8.3. Alkaline agarose gels were 1.4% agarose in 0.03 M NaOH/2 mM EDTA (23). Acrylamide gels were gradients of 5% acrylamide/0.17% bisacrylamide to 10% acrylamide/0.3% bisacrylamide with a 2.4% acrylamide/0.008% bisacrylamide sample well. The gel contained 8 M urea/0.04 M NaOAc/0.05 M Tris/2 mM EDTA, pH 8.3. The length of single-stranded DNA fragments or RNA·DNA duplex segments was determined by their mobility relative to standards of restriction fragments of Ad2 DNA. The assumed lengths of these *Hin*dIII and *Sma* I Ad2 fragments are reported in the legend to Fig. 3. The relative mobility of intact SV40 DNA in either the alkaline or neutral gel yields a length of 5000 base pairs. Recent sequence data suggest a length of approximately 5225 base pairs. The errors for the length of fragments cited in the text represent the experimental reproducibility of the relative mobilities of bands. Because the mobility of a fragment is probably affected by base sequence composition as well as length, the lengths and errors reported here cannot necessarily be directly converted into absolute numbers of nucleotides in a specific sequence.

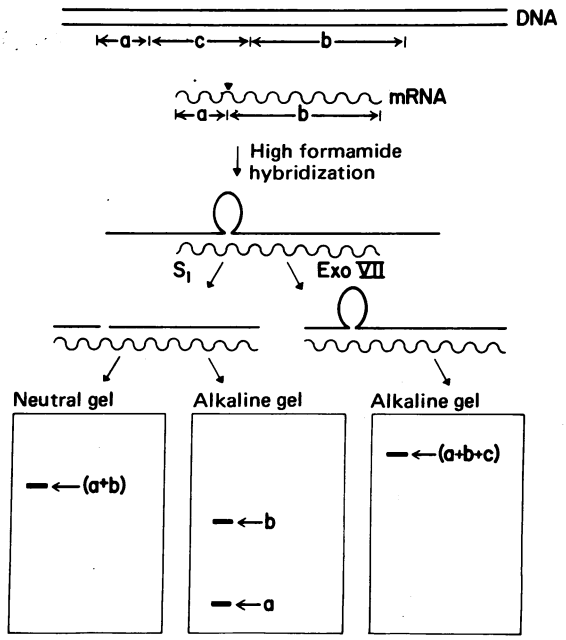

FIG. 2. Strategy for the analysis of spliced mRNA structure by gel electrophoresis of endonuclease S_1- and exonuclease VII-digested RNA·DNA hybrids.

RESULTS

The strategy for the analysis of viral mRNA structure is depicted in Fig. 2 and is as follows. Total cytoplasmic RNA is hybridized to ^{32}P-labeled viral DNA under conditions above the melting temperature (t_m) of the viral DNA but below the t_m of the RNA·DNA hybrid (20). If viral mRNAs are spliced, hybrid structures will result that are similar to those observed in the electron microscope by Berget *et al.* (11). RNA·DNA duplex is flanked by single-stranded DNA, and loops of nonhybridized single-stranded DNA result at splice points in the mRNA. When these structures are treated with the single-strand specific endonuclease S_1 (21), the single-stranded DNA is hydrolyzed, resulting in a fully duplex structure with discontinuities in the DNA at the splice points. When the S_1-treated hybrids are resolved by electrophoresis through neutral agarose gels, a band is observed migrating as expected for a duplex DNA molecule equal in length to the total mRNA (Fig. 2, lengths a + b). When these bands are excised and analyzed further by electrophoresis after denaturation, the resulting single-stranded DNA fragments observed are equal in length to the sequences spliced together to form the mRNA molecule. We refer to RNA segments transcribed from a contiguous set of DNA sequences as colinear transcripts (Fig. 2, lengths a and b). If the initial hybridization is performed using DNA digested with restriction endonucleases, patterns are produced that define the map positions of the colinear transcripts (10).

The mRNA·genome DNA hybrids are also analyzed by digestion with single-strand specific exonuclease VII of *E. coli* (ref. 22; S. Goff and P. Berg, personal communication). This exonuclease, which digests processively in both the 5′ and 3′ directions (24), removes the DNA single strands extending beyond the 5′ and 3′ ends of the mRNA but does not remove the single-stranded DNA loops that result at splice points. By analyzing the length of the exonuclease VII-resistant DNA in alkaline agarose gels, the length along the genome between the 5′ sequence and the 3′ sequence present in the mRNA is determined (Fig. 2, lengths a + b + c). Because lengths a and b

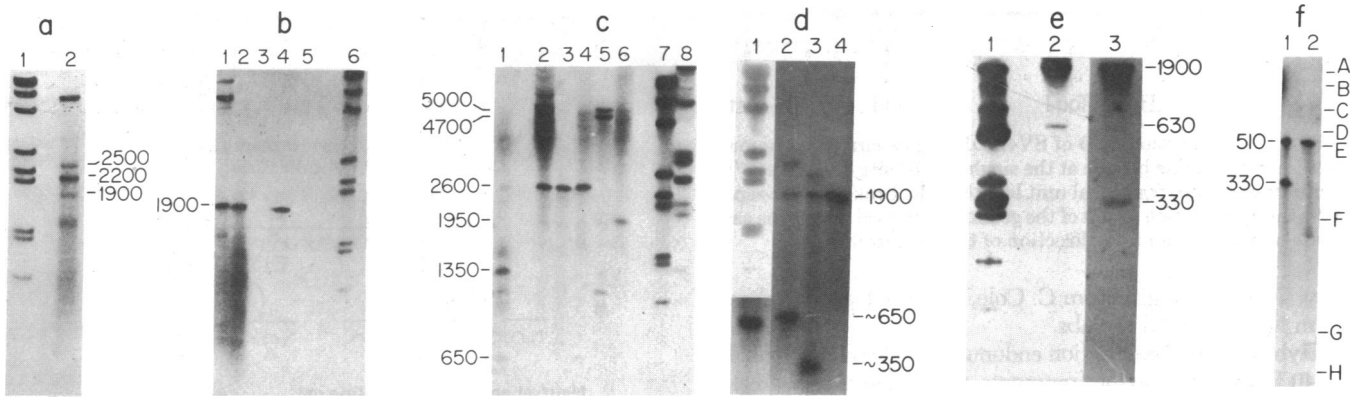

FIG. 3. Autoradiograms of gels of endonuclease S_1- and exonuclease VII-digested hybrids. (*a*) Neutral agarose gel of early cytoplasmic RNA hybridized to ^{32}P-labeled SV40 DNA cut at 0 with *Eco*RI and digested with endonuclease S_1 (track 2). Track 1, marker *Sma* I digest of Ad2 DNA. The band closest to the top of track 2 is full-length linear SV40 DNA resulting from renaturation of intact strands during quenching of the hybridization before treatment with S_1. (*b*) Alkaline agarose gel of S_1-digestion products of early RNA hybridized to: SV40 DNA cut at 0 with *Eco*RI (track 1); SV40 DNA cut at 0.14, 0.57, and 0.67 with *Bam* I, *Taq* I, and *Bgl* I (track 2); the DNA fragment from 0.67 to 0.14 (track 3); the DNA fragment from 0.14 to 0.57 (track 4); the DNA fragment from 0.57 to 0.67 (track 5) (all in the clockwise direction). Track 6, marker *Sma* I digest of Ad2 DNA. (*c*) Alkaline agarose gel of: S_1-digestion products of early RNA hybridized to SV40 DNA cut at 0.04 and 0.27 with *Pst* I (track 1). Exonuclease VII digestion products of early RNA hybridized to SV40 DNA cut at: 0 with *Eco*RI (track 2); 0 and 0.14 with *Eco*RI and *Bam* I (track 3); 0.67 with *Bgl* I (track 4); 0.57 with *Taq* I (track 5); 0.27 and 0.04 with *Pst* I (track 6). Also shown: marker digests of Ad2 DNA digested with *Sma* I (track 7) and *Hin*dIII (track 8). The lengths (kilobases) of the Ad2 *Sma* I restriction fragments are: A, 7.04; B, 6.27 (A and B run as a doublet); C, 5.22; D, 4.24; E, 2.84; F, 2.31; G, 2.21; H, 1.51; I, 1.33; J, 1.05; K, 0.63. The lengths of the Ad2 *Hin*dIII restriction fragments are: A, 8.23; B, 5.08; C, 3.33; D, 3.19; E, 3.12 (C, D, and E run as a triplet); F, 2.73; G, 2.63 (F and G run as a doublet); H, 2.21; I, 2.03; J, 1.30; K, 0.95. The bands at the top of the gel in *b* track 1 and *c* track 2 are due to a small fraction of SV40 DNA that was not linearized by *Eco*RI digestion. From the top of the gel downward they represent single-strand circles, single-strand linears, and form I DNA. (*d*) The 2500 (track 2), 2200 (track 3), and 1900 (track 4) nucleotide RNA·DNA hybrid bands cut from a preparative neutral agarose gel, similar to that shown in *a*, after hybridization to nick-translated DNA. The gel slices were melted, diluted with an equal volume of water, and layered on an alkaline agarose gel. Track 1, *Sma* I Ad2 markers. (*e*) The 2500 (track 2) and 2200 (track 3) nucleotide RNA·DNA hybrids prepared as in *d* after hybridization to *in vivo* labeled DNA denatured with alkali and layered on a 5–10% 8 M urea gradient acrylamide gel. Track 1, marker *Hae* III restriction fragments of SV40 DNA: A, 1465; B, 820; C, 550; D, 370; E, 335; F_{1a}, 315; F_{1b}, 310; F_2, 300; G, 220; H, 165 nucleotides. (*f*) S_1 products of early RNA hybridized to the nick-translated *Taq* I/*Bgl* I B fragment alkali-denatured and electrophoresed on a 5–10% 8 M urea gradient acrylamide gel (track 1). Untreated denatured *Taq* I/*Bgl* I B fragment (track 2). The positions of marker *Hin*f SV40 restriction fragments are shown: A, 1790; B, 1120; C, 760; D, 570; E, 490; F, 235; G, 110; H, 75 nucleotides.

have been determined, the genome sequence, c, between the colinear transcripts can be calculated.

As a first step in this analysis, we defined the map coordinates of the long (>1000 nucleotides) colinear transcripts present in early cytoplasmic RNA. Early SV40 cytoplasmic RNA was hybridized to ^{32}P-labeled SV40 DNA cut once in the late region at 0 with *Eco*RI. After S_1 treatment and electrophoresis on alkaline agarose gels, a major band of 1900 nucleotides was reproducibly observed (Fig. 3*b*, track 1). Often, there was a smear of DNA present on the gel below the major 1900-nucleotide band, as shown in this example. A major 1900-nucleotide band was also reproducibly observed when early RNA was hybridized to DNA cut at 0.67 (*Bgl* I), 0.14 (*Bam* I), or 0.57 (*Taq* I), or at all three of these positions (Fig. 3*b*, track 2). The same band was observed when the DNA fragment mapping from 0.57 (*Taq* I) to 0.14 (*Bam* I) was used in the hybridization (Fig. 3*b*, track 4). It should be noted that, when purified fragments were used in the hybridization, the smear of DNA below the 1900-nucleotide band was not observed. The 1900-nucleotide early colinear transcript must map from 0.54 ± 0.01 to 0.14 ± 0.01 because hybridization to DNA cut at 0.27 and 0.04 (*Pst* I) results in bands migrating at lengths of 1350 and 650 nucleotides on alkaline agarose gels (Fig. 3*c*, track 1). Hybridization to SV40 DNA cut with *Hpa* I (at positions 0.175, 0.375, and 0.73) resulted in bands migrating at the position of the 0.375–0.175 fragment and at 800 nucleotides, in agreement with this mapping position (data not shown).

When an excess of *Eco*RI-cut DNA was hybridized to early cytoplasmic RNA and the products were digested with exonuclease VII, a fragment of DNA 2600 nucleotides long was generated (Fig. 3*c*, track 2). A 2600-nucleotide band also was observed after exonuclease VII treatment of early RNA hybridized to DNA cut with *Eco*RI plus *Bam* I and to DNA cut with *Bgl* I (Fig. 3*c*, tracks 3 and 4). This exonuclease VII-resistant fragment maps from 0.67 ± 0.02 to 0.14 ± 0.01 because bands of length 1950 and 650 nucleotides are generated by this procedure when the hybridized DNA is initially cut with *Pst* I (at sites 0.27 and 0.04) (Fig. 3*c*, track 6). From these results we conclude that the early SV40 mRNAs detected by these methods must contain sequences at their 5′ ends mapping at 0.67 ± 0.02 and sequences at their 3′ ends mapping at 0.14 ± 0.01. Because the 1900-nucleotide colinear transcript maps from 0.54 ± 0.01 to 0.14 ± 0.01, we conclude that the early SV40 mRNAs must have additional sequences spliced to the 5′ end of the 1900-nucleotide colinear transcript.

To ensure that the processive digestion of single-stranded SV40 DNA by exonuclease VII had not been blocked by an inverted repeat in the DNA sequence, which occurs near position 0.67 (25), the following control experiment was performed. ^{32}P-Labeled SV40 DNA cut at 0.67 with *Bgl* I was hybridized to a 10-fold molar excess of the unlabeled SV40 DNA *Hpa* I fragment mapping from 0.175 to 0.375 and the hybridization was allowed to proceed until 80% of the fragment had renatured. The products of the hybridization were digested with exonuclease VII under the same conditions used above and the resistant fragments were analyzed by alkaline gel electrophoresis. Autoradiography revealed a faint band comigrating with full-length SV40 DNA and a dense band comigrating with the *Hpa* I fragment (data not shown). These results indicate that, under the conditions of digestion used, there are no blocks

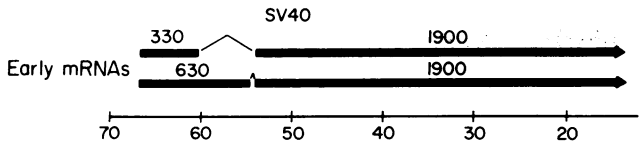

FIG. 4. Structure of the early SV40 mRNAs as determined by the data presented in Fig. 3. Heavy lines represent genome sequences included in the mRNAs and arrowheads indicate the 3′ direction. The caret-shaped symbol indicates that sequences are covalently joined in the mRNA molecule by a 3′-5′ phosphodiester bond. Numbers above the heavy lines represent lengths in nucleotides. Numbers below the narrow line indicate SV40 genome map coordinates in hundreths of a unit.

to the processive digestion of single-stranded SV40 DNA by exonuclease VII in sequences mapping counterclockwise from 0.67 to 0.375 and clockwise from 0.67 to 0.175.

To determine the nature of the sequences spliced to the 5′ end of the 1900-nucleotide colinear transcript, the following experiments were performed. SV40 DNA cut at 0 (*Eco*RI) was hybridized to early RNA and treated with S_1, and the products were run on a neutral agarose gel. Three bands were reproducibly observed at 2500 nucleotides, 2200 nucleotides, and 1900 nucleotides (Fig. 3*a*, track 2). In addition, a broad band was often observed at 1550–1650 nucleotides. To determine the sizes of the colinear transcripts from which the RNA molecules in these RNA·DNA hybrids are composed, these bands were cut from a preparative gel and analyzed by electrophoresis through an alkaline agarose gel. The 2500-nucleotide RNA·DNA hybrid gave rise to two single-stranded DNA fragments migrating at 1900 and ~650 nucleotides (Fig. 3*d*, track 2). The 2200-nucleotide band gave rise to two bands migrating at 1900 and ~350 nucleotides, and the 1900-nucleotide band yielded a single band migrating at 1900 nucleotides (Fig. 3*d*, tracks 3 and 4). Lengths of the smaller colinear transcripts were estimated more precisely on acrylamide gels (Fig. 3*e*) to be 630 and 330 nucleotides. These same lengths were observed when the initial hybridization was to DNA cut at 0.67 with *Bgl* I. Faint bands are also observed on the alkaline agarose gel migrating at 2500 and 2200 nucleotides (Fig. 3*d*, tracks 2 and 3). These are due to background transferred from the preparative neutral agarose gel, because no colinear transcripts longer than 1900 nucleotides were identified earlier (Fig. 3*b*). When similar experiments were performed on the broad 1550- to 1650-nucleotide band cut from a neutral agarose gel of S_1-treated hybrid (Fig. 3*a*), only the broad 1550–1650 band was again observed (data not shown).

Most simply interpreted, these results demonstrate that there are two early mRNAs, each of which is spliced (Fig. 4). The larger of these is a 2500-nucleotide mRNA composed of two parts. At the 5′ end of this mRNA is a colinear transcript of 630 nucleotides with its 5′ end at ~0.67. This is spliced to a 1900-nucleotide colinear transcript with its 5′ end at 0.54 ± 0.01 and its 3′ end at 0.14 ± 0.01. The smaller of the two abundant early mRNAs, a 2200-nucleotide mRNA, is similarly composed of two parts. At the 5′ end is a colinear transcript of 330 nucleotides also having its 5′ terminus at ~0.67. It is also spliced to the 1900-nucleotide colinear transcript mapping from 0.54 ± 0.01 to 0.14 ± 0.01. We interpret the 1900-nucleotide band on the neutral S_1 gel to be due to partial S_1 cutting at the splice point in these mRNAs. We do not know the origin of the broad band observed in some neutral S_1 gels migrating with a mobility of roughly 1550–1650 nucleotides or the origin of the minor bands observed in some tracks on the alkaline S_1 gels.

To test these deduced structures we analyzed the products of early cytoplasmic RNA hybridized to SV40 DNA cut at 0.57 (*Taq* I) by digestion with exonuclease VII. Hybridization of the 2500-nucleotide mRNA to *Taq* I-cut DNA should result in a circular hybrid molecule in which both ends of the hybridized early DNA strand are in duplex regions. Thus, the entire strand should be resistant to exonuclease VII digestion. Hybridization of the 2200-nucleotide mRNA to the early strand of *Taq* I-cut DNA should also result in a circular hybrid molecule. However, in this case the ends of the hybridized single-stranded DNA should not be base-paired because DNA sequences from 0.54 to 0.60 are not present in the RNA. Consequently, a total of 300 nucleotides should be removed from the 5′ and 3′ ends of the early DNA strand before hybrid region blocks further exonuclease digestion. These expected products 5000 and 4700 nucleotides long were observed as the most prominent bands on an alkaline agarose gel of exonuclease VII-digested hybrids between early RNA and *Taq* I-cut SV40 DNA (Fig. 3*c*, track 5).

The deduced structure of the early SV40 RNAs includes sequences mapping from 0.67 (*Bgl* I) to 0.57 (*Taq* I). Yet, when this isolated restriction fragment was hybridized to early RNA at 49°, no S_1-protected DNA was observed (Fig. 3*b*, track 5). The explanation for this observation is that the t_m of this restriction fragment in the hybridization buffer is 36°, ~10° lower than the calculated melting temperature (t_m) of the total genome (data not shown). When hybridization of early RNA to this restriction fragment is performed at 40° rather than at 49° and the products are treated with S_1, protection of a 330-nucleotide fragment is observed (Fig. 3*f*) as predicted by the proposed structure of the 2200-nucleotide mRNA. Protection of the full length *Taq*/*Bgl* fragment is also observed, as expected from the proposed structure of the 2500-nucleotide mRNA.

The SV40 DNA sequence near position 0.54 contains a stretch of 18 A·T base pairs (9). We were concerned that the S_1-sensitive site observed in the DNA at position 0.54 in RNA·DNA hybrid of the 2500-nucleotide mRNA might be due to cutting at this A·T(U)-rich region in a perfectly base-paired duplex, rather than the result of a splice in the 2500-nucleotide mRNA. Therefore, a control experiment was performed in which cRNA transcribed from SV40 DNA by *E. coli* RNA polymerase was hybridized to the *Hin*d II-III A restriction fragment (0.43–0.66) and the products were digested with S_1 under the conditions used in this study. The full-length restriction fragment was protected from S_1 digestion—i.e., there was no cutting at position 0.54. Therefore, the S_1-sensitive site in the DNA of the 2500-nucleotide mRNA·SV40 DNA hybrid is due to a splice in this mRNA at map position 0.54.

DISCUSSION

Fig. 4 represents the deduced structure of the early SV40 mRNAs detected in this work. The structure of these mRNAs may provide an explanation for the seemingly paradoxical observations discussed in the introduction. The small T antigen may be translated from the 2500-nucleotide mRNA. The termination codons present in all three reading frames near position 0.54 (9) in this RNA would not be present in the 2200-nucleotide early mRNA which does not contain RNA sequences from 0.60 to 0.54. Therefore, if translation is initiated at the same AUG near the 5′ termini of the 2500-nucleotide and 2200-nucleotide mRNAs, the translation products of both mRNAs would contain the same NH_2-terminal sequences. Translation of the 2500-nucleotide mRNA would be terminated near 0.54, giving rise to small T antigen. Translation of the 2200-nucleotide mRNA would continue beyond the deleted terminators to yield the large T antigen. If this were the case,

then deletions in the genome between 0.60 and 0.54 would affect the translation product of the 2500-nucleotide mRNA, the small T antigen, but would have no effect on the translation product of the 2200-nucleotide mRNA, the large T antigen. As mentioned above, mutants with large deletions in the interval between 0.54 and 0.59 induce a wild-type large T antigen but either fail to produce a detectable small T antigen or produce an altered polypeptide with a molecular weight less than 17,000 (ref. 1; Sleigh *et al.*, personal communication). This model for the translation of the two observed spliced mRNAs is also consistent with the observation that the small T antigen mRNA sediments slightly faster than the large T antigen mRNA (3).

A similar model for the structure of the early SV40 mRNAs and their translation was proposed to explain the pattern of early protein synthesis induced by SV40 deletion mutants (1). Final confirmation of this model will require determination of the NH_2-terminal sequences of large and small T antigens and a comparison of these sequences with the DNA sequence in the region of 0.67 (25).

A number of observations suggest that the organization of genetic information in the early region of the mouse papova virus (polyoma virus) may be analogous to that found in SV40. Benjamin (26) selected host-range mutants of polyoma virus that are not able to grow on 3T3 cells but are able to grow on a line of 3T3 cells transformed by polyoma virus. The prototype of this group has a deletion in a region of the genome analogous to the SV40 deletions mentioned above (27). Like the SV40 mutants, they fail to induce detectable levels of small T antigens that are observed in cells infected with wild-type polyoma virus, but produce a wild-type large T antigen (28).

If translation occurs through a splice point in the 2200-nucleotide mRNA as suggested by the model presented above, then the process by which the two colinear transcripts are joined must be precise. Furthermore, the chemistry of joining at the splice point must be a 5′-3′ phosphodiester bond. It is noteworthy that deletions in the genome throughout the interval between 0.59 and 0.54 do not affect the production of large T antigen or, therefore, of its message. This suggests that these intervening sequences do not play a role in the process by which this spliced mRNA is produced. Therefore, the process of splicing may be specified by relatively short sequences in the genome in the immediate vicinity of the sequences joined at a splice point. Evidence has been obtained that strongly suggests that the spliced late Ad2 mRNAs are generated from long initial transcripts by removal of RNA sequence between splice points (29, 30). If this is also the case for the early SV40 mRNAs, then clearly the relative abundance of these two early mRNAs would be controlled by post-transcriptional splicing. This would allow coordinate control of expression of the two early mRNAs by regulation of transcription initiation at a single promoter and yet allow another mechanism to control the relative abundance of these two messages and, hence, the relative cellular concentrations of their translation products.

We thank C. Cole for helpful discussion and for the gift of *Taq* I, S. Goff for the gift of exonuclease VII of *E. coli*, and S. Goff and S. Weissman for communication of results prior to publication. P.A.S. gratefully acknowledges a grant (VC-151A) and a Faculty Research Award from the American Cancer Society and a Cancer Center Core Grant (CA-14051). A.J.B. gratefully acknowledges a Helen Hay Whitney Foundation Postdoctoral Fellowship (F-340).

1. Crawford, L. V., Cole, C. N., Smith, A. E., Paucha, E., Tegtmeyer, P., Rundell, K. & Berg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 117–121.
2. Prives, C., Gilboa, E., Revel, M. & Winocour, E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 457–461.
3. Paucha, E., Harvey, R., Smith, R. & Smith, A. E. (1977) *INSERM, Collog.* **49,** in press.
4. Rundell, K., Collins, J. K., Tegtmeyer, P., Ozer, H. L., Lai, C. J. & Nathans, D. (1977) *J. Virol.* **21,** 636–646.
5. Simons, D. T. & Martin, M. A. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1131–1135.
6. Khoury, G., Martin, M. A., Lee, T. N. H., Danna, K. J. & Nathans, D. (1973) *J. Mol. Biol.* **78,** 377–389.
7. Sambrook, J., Sugden, B., Keller, W & Sharp, P. A. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 3711–3715.
8. Shenk, T. E., Carbon, J. & Berg, P. (1976) *J. Virol.* **18,** 664–671.
9. Thimmapaya, B. & Weissman, S. M. (1977) *Cell* **11,** 837–843.
10. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12,** 721–732.
11. Berget, S. M., Moore, C. & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3171–3175.
12. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977) *Cell* **12,** 1–8.
13. Klessig, D. F. (1977) *Cell* **12,** 9–21.
14. Aloni, Y., Dhar, R., Laub, O., Horowitz, M. & Khoury, G. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3686–3690.
15. Hsu, M.-T. & Ford, J. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4982–4985.
16. Celma, M. L., Dhar, R. & Weissman, S. M. (1977) *Nucleic Acids Res.* **4,** 2549–2559.
17. Sambrook, J., Sharp, P. A. & Keller, W. (1972) *J. Mol. Biol.* **70,** 57–71.
18. Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
19. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1184–1188.
20. Casey, J. & Davidson, N. (1977) *Nucleic Acids Res.* **4,** 1539–1552.
21. Vogt, V. M. (1973) *Eur. J. Biochem.* **33,** 192–200.
22. Chase, J. W. & Richardson, C. C. (1974) *J. Biol. Chem.* **249,** 4545–4552.
23. McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110,** 119–146.
24. Chase, J. W. & Richardson, C. C. (1974) *J. Biol. Chem.* **249,** 4553–4561.
25. Subramanian, K. N., Reddy, B. V. & Weissman, S. M. (1977) *Cell* **10,** 497–507.
26. Benjamin, T. L. (1970) *Proc. Natl. Acad. Sci. USA* **67,** 394–399.
27. Feunteun, J., Sompayrac, L., Fluck, M. & Benjamin, T. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 4169–4173.
28. Schaffhausen, B. S., Silver, J. E. & Benjamin, T. L. (1978) *Proc. Natl. Acad. Sci. USA,* **75,** 79–83.
29. Goldberg, S., Weber, J. & Darnell, J. E. (1977) *Cell* **10,** 617–621.
30. Berget, S. M., Berk, A. J., Harrison, T. & Sharp, P. A. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42,** in press.